Ensayos Bernoulli y otras cositas
(Borradores, Curso 23)

Sebastian Grynberg

15-17 de abril de 2013

Jakob Bernoulli (1654 - 1705)

En la "buena" te encontré
y en la "mala" te perdí ...

(Enrique Cadícamo)

Índice

1. Ensayos Bernoulli

Se trata de ensayos repetidos en forma independiente en los que hay sólo dos resultados posibles, usualmente denominados *"éxito"* y *"fracaso"*, cuyas probabilidades, p y $1-p$, se mantienen constantes a lo largo de todos los ensayos.

El espacio muestral de cada ensayo individual está formado por dos puntos S y F. El espacio muestral de n ensayos Bernoulli contiene 2^n puntos o secuencias de n símbolos S y F, cada punto representa un resultado posible del experimento compuesto. Como los ensayos son independientes las probabilidades se multiplican. En otras palabras, *la probabilidad de cada sucesión particular es el producto que se obtiene reemplazando los símbolos S y F por p y $1-p$, respectivamente. Así,*

$$\mathbb{P}(SSFSF\dots FFS) = pp(1-p)p(1-p)\cdots(1-p)(1-p)p.$$

Ejemplo 1.1. Si repetimos en forma independiente un experimento aleatorio y estamos interesados en la ocurrencia del evento A al que consideramos "éxito", tenemos ensayos Bernoulli con $p = \mathbb{P}(A)$. □

Modelando ensayos Bernoulli. Los ensayos Bernoulli (con probabilidad de éxito p) se describen mediante una sucesión de variables aleatorias independientes e idénticamente distribuidas $(X_i : i \in \mathbb{N})$ cada una con distribución Bernoulli(p),

$$\mathbb{P}(X_i = x_i) = p^{x_i}(1-p)^{1-x_i}, \qquad x_i \in \{0,1\}. \tag{1}$$

Esto es, $\mathbb{P}(X_i = 1) = p$ y $\mathbb{P}(X_i = 0) = 1-p$. En este contexto, $X_i = 1$ significa que *"el resultado del i-ésimo ensayo es éxito"*.

Preguntas elementales. Se pueden formular varios tipos de preguntas relacionadas con los ensayos Bernoulli. Las más sencillas son las siguientes:

(a) ¿Cuál es la cantidad total de éxitos en los primeros n ensayos?

(b) ¿En n ensayos, cuál es el número de éxitos más probable?

(c) ¿Cuánto "tiempo" hay que esperar para observar el primer éxito?

(d) ¿Cuánto "tiempo" hay que esperar para observar el k-ésimo éxito?

En lo que sigue expresaremos las preguntas (a)-(d) en términos de las variables aleatorias X_i, $i \geq 1$, que describen los ensayos Bernoulli.

La *cantidad de éxitos en los primeros n ensayos* se describe mediante la suma de las primeras variables $X_1, \dots, X_n$

$$S_n := \sum_{i=1}^{n} X_i. \tag{2}$$

La pregunta (a) interroga por la distribución de probabilidades de la variable aleatoria S_n definida en (2). Esto es, para cada $k = 0, \dots, n$, se trata de determinar cuánto valen las probabilidades $\mathbb{P}(S_n = k)$. En cambio, la pregunta (b) interroga por el valor de k que maximiza a la función de k, $\mathbb{P}(S_n = k)$.

El *tiempo de espera hasta el primer éxito* se describe mediante la variable aleatoria

$$T_1 := \text{mín}\{i \in \mathbb{N} : X_i = 1\}, \tag{3}$$

y en general, el *tiempo de espera hasta el k-ésimo éxito*, $k \geq 1$ se describe, recursivamente, mediante

$$T_k := \text{mín}\{i > T_{k-1} : X_i = 1\}. \tag{4}$$

La pregunta (c) interroga por la distribución de probabilidades de la variable T_1 definida en (3): cuánto valen las probabilidades $\mathbb{P}(T_1 = n)$, $n \in \mathbb{N}$? Finalmente, la pregunta (d) interroga por la distribución de probabilidades de las variables T_k, $k \geq 2$, definidas en (4): cuánto valen las probabilidades $\mathbb{P}(T_k = n)$, $n \geq k$?

1.1. La distribución binomial: cantidad de éxitos en n ensayos

La cantidad de éxitos puede ser $0, 1, \dots, n$. El primer problema es determinar las correspondientes probabilidades. El evento *en n ensayos resultaron k éxitos y $n - k$ fracasos*

$$\left\{(X_1, \dots, X_n) = (x_1, \dots, x_n) : \sum_{i=1}^{n} x_i = k\right\}$$

puede ocurrir de tantas formas distintas como k símbolos 1 se puedan ubicar en n lugares. En otras palabras, el evento considerado contiene $\binom{n}{k}$ puntos, cada uno de probabilidad

$$\begin{aligned}\mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i = x_i\}\right) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \\ &= p^k(1-p)^{n-k}.\end{aligned}$$

Por lo tanto,

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad 0 \leq k \leq n. \tag{5}$$

En particular, la probabilidad de que no ocurra ningún éxito en n ensayos es $(1-p)^n$ y la probabilidad de que ocurra al menos un éxito es $1 - (1-p)^n$.

La distribución de S_n, determinada en (5), se denomina *la distribución binomial de parámetros n y p* y se denota Binomial(n, p).

Nota Bene. Por definición, la distribución binomial de parámetros n y p *es la distribución de una suma de n variables aleatorias independientes cada con distribución Bernoulli de parámetro p.*

Ejemplo 1.2. Se tira un dado equilibrado 11 veces y en cada tiro se apuesta al 6, ¿cuál es la probabilidad de ganar exactamente 2 veces? Como el dado es equilibrado, la probabilidad de éxito es 1/6 y la cantidad de éxitos en 11 tiros tiene distribución Binomial $(11, 1/6)$. Por lo tanto, la probabilidad requerida es

$$\binom{11}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^9 = 0.2960\ldots$$

□

Ejemplo 1.3. Cada artículo producido por una máquina será defectuoso con probabilidad 0.1, independientemente de los demás. En una muestra de 3, ¿cuál es la probabilidad de encontrar a lo sumo un defectuoso?

Si X es la cantidad de artículos defectuosos en la muestra, entonces $X \sim \text{Binomial}(3,\, 0.1)$. En consecuencia,

$$\mathbb{P}(X \leq 1) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = \binom{3}{0}(0.1)^0(0.9)^3 + \binom{3}{1}(0.1)^1(0.9)^2 = 0.972.$$

□

Ejemplo 1.4. Un avión se mantendrá en vuelo mientras funcionen al menos el 50 % de sus motores. Si cada motor del avión en vuelo puede fallar con probabilidad $1-p$ independientemente de los demás, ¿para cuáles valores de $p \in (0,1)$ es más seguro un avión de 4 motores que uno de 2?

Como cada motor puede fallar o funcionar independientemente de los demás, la cantidad de motores que siguen funcionando es una variable aleatoria con distribución binomial. La probabilidad de que un avión de 4 motores realice un vuelo exitoso es

$$\binom{4}{2}p^2(1-p)^2 + \binom{4}{3}p^3(1-p) + \binom{4}{4}p^4 = 6p^2(1-p)^2 + 4p^3(1-p) + p^4,$$

mientras que la correspondiente probabilidad para un avión de 2 motores es

$$\binom{2}{1}p(1-p) + \binom{2}{2}p^2 = 2p(1-p) + p^2.$$

En consecuencia, el avión de 4 motores es más seguro que el de 2 si

$$6p^2(1-p)^2 + 4p^3(1-p) + p^4 > 2p(1-p) + p^2$$

lo que es equivalente a las siguientes expresiones simplificadas

$$3p^3 - 8p^2 + 7p - 2 > 0 \iff 3(p-2/3)(p-1)^2 > 0 \iff p > 2/3.$$

Por lo tanto, el avión de 4 motores es más seguro cuando la probabilidad de que cada motor se mantenga en funcionamiento es mayor que 2/3, mientras que el avión de 2 motores es más seguro cuando esa probabilidad es menor que 2/3. □

Ejemplo 1.5. Si la probabilidad de éxito es $p = 0.01$, cuántos ensayos se deben realizar para asegurar que la probabilidad de que ocurra por lo menos un éxito sea al menos 1/2?

Buscamos el menor entero n tal que $1 - (0.99)^n \geq \frac{1}{2}$, o equivalentemente $\frac{1}{2} \geq (0.99)^n$. Tomando logaritmos $-\log 2 \geq n \log(0.99)$ y despejando n resulta $n \geq -\log(2)/\log(0.99) \approx 68.96$. Por lo tanto, $n = 69$. □

1.2. Término central

De la fórmula (5) se puede ver que

$$\begin{aligned}\frac{\mathbb{P}(S_n = k)}{\mathbb{P}(S_n = k-1)} &= \frac{\binom{n}{k}p^k(1-p)^{n-k}}{\binom{n}{k-1}p^{k-1}(1-p)^{n-k+1}} = \frac{(k-1)!(n-k+1)!p}{k!(n-k)!(1-p)} \\ &= \frac{(n-k+1)p}{k(1-p)} = 1 + \frac{(n+1)p - k}{k(1-p)}. \end{aligned} \tag{6}$$

De (6) se deduce que $\mathbb{P}(S_n = k)$ crece cuando $k < (n+1)p$ y decrece cuando $k > (n+1)p$. Si $(n+1)p$ es un número entero, entonces $\mathbb{P}(S_n = (n+1)p) = \mathbb{P}(S_n = (n+1)p - 1)$. En otras palabras, la cantidad más probable de éxitos en n ensayos es $m := [(n+1)p]$. Salvo en el caso en que $m = (n+1)p$, donde también lo es $m - 1$.

Cuando $p = \frac{1}{2}$ el resultado anterior se puede observar directamente en el triángulo de Pascal: en el centro de las filas pares está el máximo. En la región central de las filas impares hay dos máximos.

Ejemplo 1.6. Se tira un dado equilibrado n veces y en cada tiro se apuesta al 6. ¿Cuál es la cantidad más probable de éxitos cuando $n = 12$? y cuando $n = 11$?

La cantidad de éxitos tiene distribución Binomial (n, p), donde $p = 1/6$. Cuando $n = 12$, $(n+1)p = 13/6 = 2.16...$ y entonces la cantidad más probable de éxitos es $m = 2$. Cuando $n = 11$, $(n+1)p = 2$ y entonces la cantidad más probable de éxitos es $m = 1$ o $m = 2$. □

1.3. La distribución geométrica: tiempo de espera hasta el primer éxito

El tiempo que hay que esperar para observar el primer éxito en una sucesión de ensayos Bernoulli puede ser $n = 1, 2, \dots$. El evento $T_1 = 1$ significa que se obtuvo éxito en el primer ensayo y tiene probabilidad p. Para cada $n \geq 2$, el evento $T_1 = n$ significa que en los primeros $n-1$ ensayos se obtuvieron fracasos y que en el n-ésimo se obtuvo éxito, lo que tiene probabilidad $(1-p)^{n-1}p$. Por lo tanto, la distribución de T_1 es

$$\mathbb{P}(T_1 = n) = (1-p)^{n-1}p, \qquad n \in \mathbb{N}. \tag{7}$$

El evento $T_1 > n$ significa que los primeros n ensayos de la sucesión resultaron fracaso. Por lo tanto,

$$\mathbb{P}(T_1 > n) = (1-p)^n, \qquad n \geq 1. \tag{8}$$

La distribución de T_1 se denomina *distribución geométrica de parámetro* p y se designa mediante Geométrica(p).

Ejemplo 1.7. Se arroja repetidamente un dado equilibrado. ¿Cuál es la probabilidad de que el primer 6 aparezca antes del quinto tiro?. La probabilidad de obtener 6 es 1/6 y la cantidad de tiros hasta obtener el primer as tiene distribución Geométrica(1/6). Por lo tanto, la probabilidad requerida es

$$1/6 + (5/6)(1/6) + (5/6)^2(1/6) + (5/6)^3(1/6) = (1/6)\left(\frac{1-(5/6)^4}{1-(5/6)}\right) = 1 - (5/6)^4 = 0.5177\dots$$

□

Ejemplo 1.8 (Ocurrencias casi seguras)**.** Si al realizarse un experimento aleatorio un evento A tiene probabilidad positiva de ocurrir, entonces en una sucesión de experimentos independientes el evento A ocurrirá *casi seguramente*.

En efecto, el tiempo de espera hasta que ocurra el evento A es una variable aleatoria T_A con distribución geométrica de parámetro $p = \mathbb{P}(A)$. Si se observa que

$$\{T_A > 1\} \supseteq \{T_A > 2\} \supseteq \{T_A > 3\} \supseteq \cdots$$

y que

$$\{T_A = \infty\} = \bigcap_{n\geq 1} \{T_A > n\}$$

y se usa la propiedad de continuidad de $\mathbb{P}$, se obtiene que

$$\mathbb{P}(T_A = \infty) = P\left(\bigcap_{n\geq 1} \{T_A > n\}\right) = \lim_{n\to\infty} \mathbb{P}(T_A > n) = \lim_{n\to\infty} (1-p)^n = 0.$$

Por lo tanto, $\mathbb{P}(T_A < \infty) = 1$. $\square$

Pérdida de memoria

La variable aleatoria, T, con distribución geométrica de parámetro p tiene la propiedad de *pérdida de memoria*,

$$\mathbb{P}(T > n + m | T > n) = \mathbb{P}(T > m) \qquad n, m \in \mathbb{N} \tag{9}$$

La identidad (9) se obtiene de (8) y de la fórmula de probabilidad condicional:

$$\begin{aligned} \mathbb{P}(T > n + m | T > n) &= \frac{\mathbb{P}(T > n + m, T > n)}{\mathbb{P}(T > n)} \\ &= \frac{\mathbb{P}(T > n + m)}{\mathbb{P}(T > n)} = \frac{(1-p)^{n+m}}{(1-p)^n} \\ &= (1-p)^m = \mathbb{P}(T > m). \end{aligned}$$

De hecho, la propiedad de pérdida de memoria definida en (9) caracteriza a la distribución geométrica.

Teorema 1.9. Si T es una variable aleatoria a valores en $\mathbb{N}$ con la propiedad de pérdida de memoria, entonces $T \sim$ Geométrica(p), donde $p = \mathbb{P}(T = 1)$.

Demostración. Sea $G(n) := \mathbb{P}(T > n)$. Si T pierde memoria, tenemos que

$$G(n + m) = G(n)G(m) \tag{10}$$

De (10) sigue que $G(2) = G(1)G(1) = G(1)^2$, $G(3) = G(2)G(1) = G(1)^3$ y en general $G(n) = G(1)^n$ cualquiera sea $n \in \mathbb{N}$. En otros términos, la distribución de T es tal que

$$\mathbb{P}(T > n) = G(1)^n.$$

Por lo tanto,

$$\mathbb{P}(T = n) = \mathbb{P}(T > n - 1) - \mathbb{P}(T > n) = G(1)^{n-1} - G(1)^n = G(1)^{n-1}(1 - G(1)).$$

$\square$

1.4. La distribución Pascal: tiempo de espera hasta el k-ésimo éxito

Si se quieren observar k-éxitos en una sucesión de ensayos Bernoulli lo mínimo que se debe esperar es k ensayos. ¿Cuándo ocurre el evento $T_k = n$, $n \geq k$? El n-ésimo ensayo debe ser éxito y en los $n-1$ ensayos anteriores deben ocurrir exactamente $k-1$ éxitos. Hay $\binom{n-1}{k-1}$ formas distintas de ubicar $k-1$ símbolos 1 en $n-1$ lugares. Por lo tanto,

$$\mathbb{P}(T_k = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k} \qquad n \geq k. \tag{11}$$

La distribución de T_k se denomina *distribución Pascal de parámetros k y p* y se designa mediante Pascal(k, p).

La distribución Pascal de parámetros k y p es la distribución de una suma de k variables aleatorias independientes cada una con ley Geométrica(p). Lo cual es intuitivamente claro si se piensa en el modo que arribamos a su definición.

En efecto, definiendo $T_0 := 0$ vale que

$$T_k = \sum_{i=1}^{k} (T_i - T_{i-1}).$$

Basta ver que para cada $i = 1, \dots, k$ las diferencias $T_i - T_{i-1}$ son independientes y todas se distribuyen como $T_1 \sim$ Geométrica(p). De acuerdo con la regla del producto

$$\begin{aligned} \mathbb{P}\left(\cap_{i=1}^{k}\{T_i - T_{i-1} = m_i\}\right) &= \mathbb{P}(T_1 = m_1) \\ &\quad \times \prod_{i=2}^{n-1} \mathbb{P}\left(T_i - T_{i-1} = m_i | \cap_{j=1}^{i-1} \{T_j - T_{j-1} = m_j\}\right). \end{aligned} \tag{12}$$

Si se sabe que $T_1 = m_1, \dots, T_{i-1} - T_{i-2} = m_{i-1}$, entonces el evento $T_i - T_{i-1} = m_i$ depende las variables aleatorias $X_{\sum_{j=1}^{i-1} m_j + 1}, \dots, X_{\sum_{j=1}^{i} m_j}$ y equivale a decir que las primeras $m_i - 1$ de esas variables valen 0 y la última vale 1. En consecuencia,

$$\mathbb{P}\left(T_i - T_{i-1} = m_i | \cap_{j=1}^{i-1} \{T_j - T_{j-1} = m_j\}\right) = (1-p)^{m_i - 1} p. \tag{13}$$

De (12) y (13) se deduce que

$$\mathbb{P}\left(\cap_{i=1}^{k}\{T_i - T_{i-1} = m_i\}\right) = \prod_{i=1}^{k} (1-p)^{m_i - 1} p. \tag{14}$$

De la factorización (14) se deduce que $T_1, T_2 - T_1, \dots, T_k - T_{k-1}$ son independientes y que cada una tiene distribución geométrica de parámetro p. □

Ejemplo 1.10. Lucas y Monk disputan la final de un campeonato de ajedrez. El primero que gane 6 partidas (no hay tablas) resulta ganador. La probabilidad de que Lucas gane cada partida es 3/4. ¿Cuál es la probabilidad de que Lucas gane el campeonato en la novena partida? La cantidad de partidas que deben jugarse hasta que Lucas gane el campeonato tiene distribución Pascal$(6, 3/4)$. Por lo tanto, la probabilidad requerida es

$$\binom{8}{5} \left(\frac{3}{4}\right)^6 \left(\frac{1}{4}\right)^3 = 0.1557\dots$$

□

Ejemplo 1.11. En una calle hay tres parquímetros desocupados. Se estima que en los próximos 10 minutos pasarán 6 coches por esa calle y, en media, el 80 % tendrá que estacionarse en alguno de ellos. Calcular la probabilidad de que los tres parquímetros sean ocupados en los próximos 10 minutos.

La probabilidad requerida es la probabilidad de que la cantidad, N, de ensayos hasta el tercer éxito sea menor o igual que 6. Como N tiene distribución Pascal(3, 0.8) resulta que

$$\begin{aligned}\mathbb{P}(N \leq 6) &= \sum_{n=3}^{6} \mathbb{P}(N = n) = \sum_{n=3}^{6} \binom{n-1}{2} (0.8)^3 (0.2)^{n-3} \\ &= (0.8)^3 \left[\binom{2}{2}(0.2)^0 + \binom{3}{2}(0.2)^1 + \binom{4}{2}(0.2)^2 + \binom{5}{2}(0.2)^3 \right] \\ &= (0.8)^3 \left[1 + 3(0.2) + 6(0.2)^2 + 10(0.2)^3 \right] \\ &= 0.983\ldots\end{aligned}$$

Notar que una forma alternativa de obtener el mismo resultado es sumar las probabilidades de observar $3, 4, 5, 6$ éxitos en 6 ensayos Bernoulli. □

Relación entre las distribuciones Binomial y Pascal. Sean $S_n \sim$ Binomial(n, p) y $T_k \sim$ Pascal(k, p). Vale que

$$\mathbb{P}(S_n \geq k) = \mathbb{P}(T_k \leq n). \tag{15}$$

En efecto, decir que en n ensayos Bernoulli ocurren por lo menos k éxitos es lo mismo que decir que el tiempo de espera hasta observar el k-ésimo éxito no supera a n. □

1.5. La distribución multinomial

La distribución binomial se puede generalizar al caso de n ensayos independientes donde cada ensayo puede tomar uno de varios resultados. Sean $1, 2, \ldots, r$ los resultados posibles de cada ensayo y supongamos que para cada $k \in \{1, 2, \ldots, r\}$ la probabilidad p_k de observar el valor k se mantiene constante a lo largo de los ensayos. La pregunta es: ¿Cuántas veces ocurre cada uno de los resultados en los primeros n ensayos?

Consideramos una sucesión $X_1, X_2, \ldots$ de variables aleatorias independientes e idénticamente distribuidas a valores $\{1, 2, \ldots, r\}$ tal que $\mathbb{P}(X_i = k) = p_k$. Fijado n, para cada $k = 1, \ldots, r$ definimos la variables $M_k = \sum_{i=1}^{n} \mathbf{1}\{X_i = k\}$. La variable M_k cuenta la cantidad de veces que ocurre el resultado k en n ensayos. La probabilidad de que en n ensayos el resultado 1 ocurra m_1 veces, el resultado 2 ocurra m_2 veces, etc. es

$$\mathbb{P}(M_1 = m_1, M_2 = m_2, \ldots, M_r = m_r) = \frac{n!}{m_1! m_2! \cdots m_r!} p_1^{m_1} p_2^{m_2} \cdots p_r^{m_r}, \tag{16}$$

donde los m_k son enteros no negativos sujetos a la condición $m_1 + m_2 + \cdots + m_r = n$.

Si $r = 2$, entonces (16) se reduce a la distribución Binomial con $p_1 = p$, $p_2 = 1 - p$, $k_1 = k$ y $k_2 = n - k$.

1.6. ☢ Miscelánea de ejemplos

Observación 1.12 (Desarrollo de Taylor)**.** Para todo $x \in (0,1)$ vale que

$$\frac{1}{(1-x)^{k+1}} = \sum_{n\geq 0} \binom{n+k}{k} x^n. \tag{17}$$

La identidad (17) se obtiene desarrollando la función $h(x) = (1-x)^{-(k+1)}$ en serie de Taylor alrededor del 0: observando que $h^{(n)}(0) = (k+1)(k+2)\cdots(k+n)$, se obtiene que $\frac{h^{(n)}(0)}{n!} = \binom{n+k}{k}$. □

Ejemplo 1.13 (Variable compuesta)**.** Sean $N_1; X_1, X_2, \dots$ una sucesión de variables aleatorias independientes. Supongamos que $N_1 \sim$ Geométrica(p_1) y que $X_i \sim$ Bernoulli(p_2), $i \geq 1$. Entonces,

$$N_2 = \sum_{i=1}^{N_1 - 1} X_i \sim \text{ Geométrica}\left(\frac{p_1}{p_1 + p_2(1-p_1)}\right) - 1. \tag{18}$$

Por definición $N_2|N_1 = n \sim$ Binomial$(n-1, p_2)$. Aplicando la fórmula de probabilidad total obtenemos

$$\begin{aligned}
\mathbb{P}(N_2 = k) &= \sum_{n\geq 1} \mathbb{P}(N_2 = k|N_1 = n)\mathbb{P}(N_1 = n) \\
&= \sum_{n\geq k+1} \binom{n-1}{k} p_2^k (1-p_2)^{n-1-k}(1-p_1)^{n-1} p_1 \\
&= \sum_{m\geq 0} \binom{m+k}{k} p_2^k (1-p_2)^m (1-p_1)^{m+k} p_1 \\
&= (p_2(1-p_1))^k\, p_1 \sum_{m\geq 0} \binom{n+k}{k} [(1-p_1)(1-p_2)]^m.
\end{aligned} \tag{19}$$

Usando (17) vemos que

$$\begin{aligned}
\sum_{m\geq 0} \binom{m+k}{k} [(1-p_1)(1-p_2)]^m &= \frac{1}{(1-(1-p_1)(1-p_2))^{k+1}} \\
&= \frac{1}{(p_1 + p_2(1-p_1))^{k+1}}.
\end{aligned} \tag{20}$$

Combinando (19) y (20) obtenemos que

$$\mathbb{P}(N_2 = k) = \frac{(p_2(1-p_1))^k\, p_1}{(p_1 + p_2(1-p_1))^{k+1}} = \left(\frac{p_2(1-p_1)}{p_1 + p_2(1-p_1)}\right)^k \left(\frac{p_1}{p_1 + p_2(1-p_1)}\right). \tag{21}$$

□

Ejemplo 1.14 (Rachas)**.** Para cada número entero $m > 1$ sea Y_m la cantidad de ensayos Bernoulli(p) que se deben realizar hasta obtener por primera vez una racha de m *éxitos* seguidos. En lo que sigue vamos a calcular $E[Y_m]$ mediante condicionales. Para ello introducimos una variable aleatoria auxiliar N que cuenta la cantidad de ensayos que deben realizarse hasta obtener por primera vez un *fracaso* y usaremos la identidad $\mathbb{E}[Y_m] = \mathbb{E}[\mathbb{E}[Y_m|N]]$.

Observando que

$$Y_m|N = n \sim \begin{cases} n + Y_m & \text{si } n \leq m, \\ m & \text{si } n > m, \end{cases}$$

obtenemos la expresión de la función de regresión

$$\varphi(n) = \mathbb{E}[Y_m|N = n] = \begin{cases} n + \mathbb{E}[Y_m] & \text{si } n \leq m, \\ m & \text{si } n > m. \end{cases}$$

En consecuencia, $\mathbb{E}[Y_m|N] = N\mathbf{1}\{N \leq m\} + \mathbb{E}[Y_m]\mathbf{1}\{N \leq m\} + m\mathbf{1}\{N > m\}$, de donde se deduce que $\mathbb{E}[Y_m] = \mathbb{E}[N\mathbf{1}\{N \leq m\}] + \mathbb{E}[Y_m]\mathbb{P}(N \leq m) + m\mathbb{P}(N > m)$. Equivalentemente,

$$\mathbb{E}[Y_m] = \frac{\mathbb{E}[N\mathbf{1}\{N \leq m\}]}{\mathbb{P}(N > m)} + m. \tag{22}$$

Debido a que $N\mathbf{1}\{N \leq m\} = N - N\mathbf{1}\{N > m\}$ el primer término del lado derecho de la igualdad (22) se puede expresar de siguiente forma

$$\begin{aligned} \frac{\mathbb{E}[N\mathbf{1}\{N \leq m\}]}{\mathbb{P}(N > m)} &= \frac{\mathbb{E}[N] - \mathbb{E}[N\mathbf{1}\{N > m\}]}{\mathbb{P}(N > m)} = \frac{\mathbb{E}[N]}{\mathbb{P}(N > m)} - \mathbb{E}[N|N > m] \\ &= \frac{\mathbb{E}[N]}{\mathbb{P}(N > m)} - \mathbb{E}[N] - m. \end{aligned} \tag{23}$$

La última igualdad se deduce de la propiedad de pérdida de memoria de la distribución Geométrica. De $N|N > m \sim m + N$, resulta que $\mathbb{E}[N|N > m] = m + \mathbb{E}[N]$.

Combinando (22) y (23) obtenemos

$$\mathbb{E}[Y_m] = \frac{\mathbb{E}[N]}{\mathbb{P}(N > m)} - \mathbb{E}[N] = \frac{\mathbb{E}[N]\mathbb{P}(N \leq m)}{\mathbb{P}(N > m)} = \frac{1 - p^m}{(1 - p)p^m}. \tag{24}$$

$\square$

Ejemplo 1.15 (Coleccionista I)**.** Sea M una variable aleatoria a valores $1, 2, \ldots, m$. Sea $(M_n : n \in \mathbb{N})$ una sucesión de variables aleatorias independientes tal que $M_n \sim M$ para todo $n \in \mathbb{N}$. Sea $K = \text{mín}\{n \geq m : \{M_1, \ldots, M_n\} = \{1, 2, \ldots, m\}\}$ el tamaño de muestra mínimo que se necesita para "coleccionar" todos los valores $1, 2, \ldots, m$. En lo que sigue vamos a calcular $\mathbb{E}[K]$ mediante condicionales. Introducimos un elemento aleatorio C que indica el orden en que se obtuvieron los valores $1, 2, \ldots, m$ y usamos la identidad $\mathbb{E}[K] = \mathbb{E}[\mathbb{E}[K|C]]$.

Sea $S(m)$ al conjunto de todas las permutaciones de los números $1, 2, \ldots, m$. Para cada permutación $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_m) \in S(m)$ vale que:

$$\mathbb{P}(C = \sigma) = \prod_{k=1}^{m-1} \frac{\mathbb{P}(M = \sigma_k)}{\sum_{i=k}^{m} \mathbb{P}(M = \sigma_i)}.$$

Por otra parte

$$K|C = \sigma \sim 1 + \sum_{k=1}^{m-1} N(\sigma_i : 1 \leq 1 \leq k),$$

donde $N(\sigma_i : 1 \leq i \leq k) \sim$ Geométrica$\big(\sum_{i=k+1}^m \mathbb{P}(M = \sigma_i)\big)$. Por lo tanto,

$$\begin{aligned}
\mathbb{E}[K] &= \sum_{\sigma \in S(m)} \mathbb{E}[K|C = \sigma]\mathbb{P}(C = \sigma) \\
&= \sum_{\sigma \in S(m)} \left(1 + \sum_{k=1}^{m-1} \frac{1}{\sum_{i=k+1}^m \mathbb{P}(M = \sigma_i)}\right) \prod_{k=1}^{m-1} \frac{\mathbb{P}(M = \sigma_k)}{\sum_{i=k}^m \mathbb{P}(M = \sigma_i)}. \qquad (25)
\end{aligned}$$

En el caso particular en que $\mathbb{P}(M = i) = 1/m$ para todo $i \in \{1, 2, \dots, m\}$ tenemos que

$$\begin{aligned}
\mathbb{E}[K] &= \sum_{\sigma \in S(m)} \left(1 + \sum_{k=1}^{m-1} \frac{1}{\sum_{i=k+1}^m 1/m}\right) \prod_{k=1}^{m-1} \frac{1/m}{\sum_{i=k}^m 1/m} \\
&= m! \left(1 + \sum_{k=1}^{m-1} \frac{1}{\sum_{i=k+1}^m 1/m}\right) \frac{1}{m!} = \sum_{k=0}^{m-1} \frac{1}{\sum_{i=k+1}^m 1/m} = m \sum_{i=1}^m \frac{1}{i}. \qquad (26)
\end{aligned}$$

□

Ejemplo 1.16 (Coleccionista II)**.** Sea $X_1, X_2, \dots$ una sucesión de variables aleatorias independientes e idénticamente distribuidas a valores $1, 2, \dots, r$. Sea $N_r = \text{mín}\{n \geq 1 : X_n = r\}$. Para cada $i = 1, \dots, r-1$ sea $M_i = \sum_{n=1}^{N_r - 1} \mathbf{1}\{X_n = i\}$. Queremos hallar la función de probabilidad de M_i.

Por definición $N_r \sim$ Geométrica(p_r) y $M_i | N_r = n \sim$ Binomial$\big(n-1, p_i(1-p_r)^{-1}\big)$. De acuerdo con el Ejemplo 1.13 tenemos que

$$M_i \sim \text{ Geométrica}\left(\frac{p_r}{p_r + p_i(1-p_r)^{-1}(1-p_r)}\right) - 1 = \text{ Geométrica}\left(\frac{p_r}{p_r + p_i}\right) - 1.$$

En particular, $\mathbb{E}[M_i] = p_i/p_r$ y $\mathbb{V}(M_i) = p_i(p_r + p_i)/p_r^2$. □

2. La distribución de Poisson

2.1. Motivación: Aproximación de Poisson de la distribución binomial

En diversas aplicaciones tenemos que tratar con ensayos Bernoulli donde, para decirlo de algún modo, n es grande y p es pequeño, mientras que el producto $\lambda = np$ es moderado. En tales casos conviene usar una aproximación de las probabilidades $\mathbb{P}(S_n = k)$, donde $S_n \sim$Binomial(n, p) y $p = \lambda/n$. Para $k = 0$ tenemos

$$\mathbb{P}(S_n = 0) = (1-p)^n = \left(1 - \frac{\lambda}{n}\right)^n. \qquad (27)$$

Tomando logaritmos y usando el desarrollo de Taylor,

$$\log(1-t) = -t - \frac{1}{2}t^2 - \frac{1}{3}t^3 - \frac{1}{4}t^4 - \cdots,$$

se obtiene

$$\log \mathbb{P}(S_n = 0) = n \log\left(1 - \frac{\lambda}{n}\right) = -\lambda - \frac{\lambda^2}{2n} - \cdots \qquad (28)$$

En consecuencia, para n grande se tiene que

$$\mathbb{P}(S_n = 0) \approx e^{-\lambda}, \tag{29}$$

donde el signo $\approx$ se usa para indicar una igualdad aproximada (en este caso de orden de magnitud $1/n$). Más aún, usando la identidad (6) se puede ver que para cada k fijo y n suficientemente grande

$$\frac{\mathbb{P}(S_n = k)}{\mathbb{P}(S_n = k-1)} = \frac{(n-k+1)p}{k(1-p)} \approx \frac{\lambda}{k}. \tag{30}$$

Recursivamente se concluye que

$$\begin{aligned} \mathbb{P}(S_n = 1) &\approx \lambda \cdot \mathbb{P}(S_n = 0) \approx \lambda e^{-\lambda}, \\ \mathbb{P}(S_n = 2) &\approx \frac{\lambda}{2} \cdot \mathbb{P}(S_n = 1) \approx \frac{\lambda^2}{2} e^{-\lambda}, \end{aligned}$$

y en general

$$\mathbb{P}(S_n = k) \approx \frac{\lambda^k}{k!} e^{-\lambda}. \tag{31}$$

La igualdad aproximada (31) se llama *la aproximación de Poisson de la distribución binomial*.

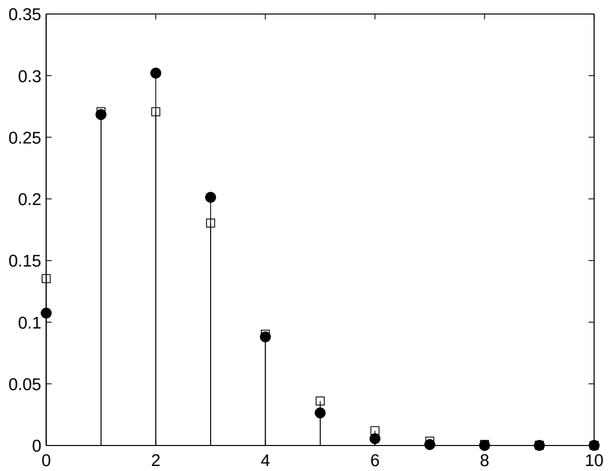

Figura 1: Comparación. Funciones de probabilidad de las distribuciones Binomial$(10, 1/5)$ (bolita negra) y Poisson(2) (cuadradillo vacío).

Otro modo de obtener el mismo resultado.

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \sim \frac{1}{k!} \left(\frac{np}{1-p}\right)^k (1-p)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}.$$

Ejemplo 2.1 (Artículos defectuosos)**.** Una industria produce tornillos. Supongamos que la probabilidad de que un tornillo resulte defectuoso sea $p = 0.015$, entonces la probabilidad de que una caja de 100 tornillos no contenga ninguno defectuoso es $(0.985)^{100} = 0.2206...$ La aproximación de Poisson es $e^{-1.5} = 0.2231...$ y es suficientemente próxima para la mayoría de los propósitos prácticos. Si se pregunta: Cuántos tornillos debería contener la caja para que la probabilidad de encontrar al menos 100 tornillos sin defectos sea 0.8 o mejor? Si $100 + x$ es el número buscado, entonces x es un número pequeño. Para aplicar la aproximación de Poisson para $n = 100 + x$ ensayos debemos poner $\lambda = np$, pero np es aproximadamente $100p = 1.5$. Buscamos el menor entero x para el cual

$$e^{-1.5}\left(1 + \frac{1.5}{1} + \cdots \frac{(1.5)^x}{x!}\right) \geq 0.8 \tag{32}$$

Para $x = 1$ el valor del lado izquierdo de la inecuación (32) es aproximadamente 0.558, para $x = 2$ es aproximadamente 0.809. Por lo tanto, la aproximación de Poisson permite concluir que se necesitan 102 tornillos. En realidad la probabilidad de encontrar al menos 100 tornillos sin defectos en una caja de 102 es $0.8022\ldots$. □

2.2. La distribución Poisson

Sea $\lambda > 0$. Una variable aleatoria N tiene distribución Poisson(λ) si sus posibles valores son los enteros no negativos y si

$$\mathbb{P}(N = n) = e^{-\lambda}\frac{\lambda^n}{n!}, \qquad n = 0, 1, \ldots \tag{33}$$

Media y varianza. Usando el desarrollo de Taylor de la función exponencial $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ se demuestra que $\mathbb{E}[N] = \lambda$ y $\mathbb{V}(N) = \lambda$.

Aditividad. El rasgo más importante de la distribución Poisson es su aditividad.

Teorema 2.2 (Aditividad)**.** *Si N_1 y N_2 son variables aleatorias independientes con distribución Poisson de medias λ_1 y λ_2, respectivamente. Entonces,*

$$N_1 + N_2 \sim Poisson(\lambda_1 + \lambda_2).$$

Demostración.

$$\begin{aligned}
\mathbb{P}(N_1 + N_2 = n) &= \sum_{m=0}^{n} \mathbb{P}(N_1 = m, N_2 = n - m) = \sum_{m=0}^{n} \mathbb{P}(N_1 = m)\mathbb{P}(N_2 = n - m) \\
&= \sum_{m=0}^{n} e^{-\lambda_1}\frac{\lambda_1^m}{m!}e^{-\lambda_2}\frac{\lambda_2^{n-m}}{(n-m)!} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{m=0}^{n}\binom{n}{m}\lambda_1^m\lambda_2^{n-m} \\
&= e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1 + \lambda_2)^n}{n!}.
\end{aligned}$$

□

Nota Bene. El resultado del Teorema 2.2 se extiende por inducción a la suma de una cantidad finita de variables aleatorias independientes con distribución Poisson.

Teorema 2.3 (Competencia)**.** *Sean $N_1, N_2, \dots, N_m$ variables aleatorias independientes, cada N_j con distribución Poisson de media λ_j, respectivamente. Sea $S = N_1 + \cdots + N_m$. Entonces, para cada $n \geq 1$ vale que*

$$(N_1, N_2, \dots, N_m)|S = n \sim Multinomial\left(n, \frac{\lambda_1}{\lambda}, \frac{\lambda_2}{\lambda}, \dots, \frac{\lambda_m}{\lambda}\right),$$

donde $\lambda = \sum_j \lambda_j$. En particular,

$$\mathbb{P}(N_j = 1|S = 1) = \frac{\lambda_j}{\lambda}.$$

Demostración. La suma $S = N_1 + \cdots + N_m$ tiene distribución Poisson de media $\lambda = \sum_j \lambda_j$; y entonces siempre que $n_1 + \cdots + n_m = n$,

$$\begin{aligned}
\mathbb{P}(N_1 = n_1, \dots, N_m = n_m|S = n) &= \frac{\mathbb{P}(N_1 = n_1, \dots, N_m = n_m)}{\mathbb{P}(S = n)} \\
&= \prod_j \left(e^{-\lambda_j} \frac{\lambda_j^{n_j}}{n_j!}\right) \Bigg/ \left(e^{-\lambda} \frac{\lambda^n}{n!}\right) \\
&= \frac{n!}{n_1! n_2! \cdots n_m!} \prod_j \left(\frac{\lambda_j}{\lambda}\right)^{n_j}.
\end{aligned}$$

□

Nota Bene. En el caso particular $n = 2$, el resultado del Teorema 2.3 se reduce a que, si N_1 y N_2 son variables aleatorias independientes con distribución Poisson de medias λ_1 y λ_2, respectivamente, entonces, dado que $N_1 + N_2 = n$, la distribución condicional de N_1 es Binomial(n, p), donde $p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$. □

Teorema 2.4 (Adelgazamiento)**.** *Sea N una variable aleatoria Poisson de media λ. Sea M una variable aleatoria tal que*

$$M|N = n \sim Binomial(n, p).$$

Entonces, M y $N - M$ son variables aleatorias independientes con distribución Poisson de medias $p\lambda$ y $(1 - p)\lambda$, respectivamente.

Demostración. Sean $m, k \geq 0$

$$\begin{aligned}
\mathbb{P}(M = m, N - M = k) &= \mathbb{P}(M = m, N - M = k|N = m + k)\mathbb{P}(N = m + k) \\
&= \mathbb{P}(M = m|N = m + k)\mathbb{P}(N = m + k) \\
&= \left(\binom{m+k}{m} p^m (1 - p)^k\right) e^{-\lambda} \frac{\lambda^{m+k}}{(m + k)!} \\
&= \left(e^{-p\lambda} \frac{(p\lambda)^m}{m!}\right) \left(e^{-(1-p)\lambda} \frac{((1 - p)\lambda)^k}{k!}\right).
\end{aligned}$$

□

Ejercicios adicionales

1. Sea N una variable aleatoria con distribución Poisson de media λ. Mostrar que

$$\mathbb{P}(N = n) = \frac{\lambda}{n}\mathbb{P}(N = n - 1), \qquad n = 1, 2, \dots$$

Usar ese resultado para encontrar el valor de n para el cual $\mathbb{P}(N = n)$ es maximal.

2. Se lanza una moneda una cantidad aleatoria N de veces, donde N tiene distribución Poisson. Sean N_1 y N_2 la cantidad de total de caras y de cecas observadas, respectivamente. Mostrar que las variables aleatorias N_1 y N_2 son independientes y que tienen distribución Poisson.

3. Sea $X_1, X_2, \dots$ una sucesión de variables aleatorias independientes, cada una con distribución Bernoulli(p). Para cada $n \geq 1$ se define $S_n := \sum_{i=1}^{n} X_i$. Por convención, $S_0 := 0$. Sea N una variable aleatoria con distribución Poisson(λ). Mostrar que $S_N \sim$ Poisson($p\lambda$).

2.3. La aproximación Poisson. (Técnica de acoplamiento)

En lo que sigue mostraremos que cuando se consideran una gran cantidad de eventos independientes y cada uno de ellos tiene una probabilidad muy pequeña de ocurrir, la cantidad de tales eventos que realmente ocurre tiene una distribución "cercana" a la distribución Poisson.

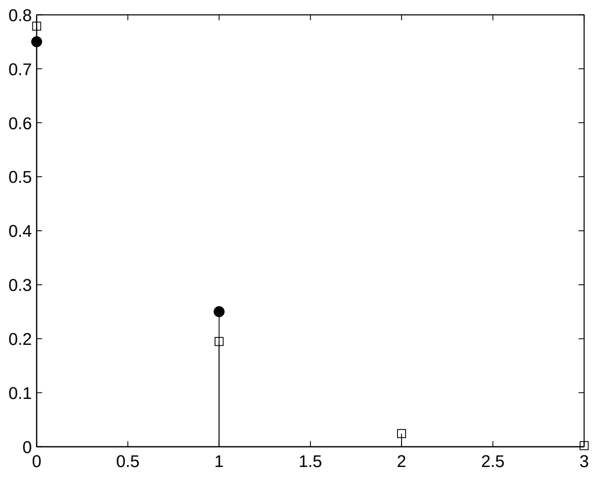

Figura 2: Comparación de las funciones de probabilidad de las distribuciones Bernoulli(1/4) (bolita negra) y Poisson(1/4) (cuadradillo vacío)

Construcción conjunta de variables Bernoulli y Poisson (Acoplamiento).

Para cada $p \in [0,1]$ dividimos el intervalo $[0,1)$ en dos intervalos

$$I_0(p) = [0, 1-p), \qquad I_1(p) = [1-p, 1) \tag{34}$$

y en la sucesión de intervalos

$$J_0(p) = [0, e^{-p}), \qquad J_k(p) = \left[\sum_{j=0}^{k-1} e^{-p}\frac{p^k}{k!}, \sum_{j=0}^{k} e^{-p}\frac{p^k}{k!}\right), \qquad k = 1, 2, \dots. \tag{35}$$

Consideramos una variable aleatoria U con distribución $\mathcal{U}[0,1)$ y construimos dos variables aleatorias V y W con distribuciones Bernoulli(p) y Poisson(p), respectivamente:

$$V := \mathbf{1}\{U \in I_1(p)\}, \qquad W := \sum_{k=0}^{\infty} k\mathbf{1}\{U \in J_k(p)\}. \tag{36}$$

De la desigualdad $1-p \leq e^{-p}$ resulta que $I_0(p) \subset J_0(p)$ y que $J_1(p) \subset I_1(p)$. En consecuencia, $V = W \iff U \in I_0(p) \cup J_1(p)$. Por ende,

$$\mathbb{P}(V = W) = \mathbb{P}(U \in I_0(p) \cup J_1(p)) = 1 - p + e^{-p}p, \tag{37}$$

y en consecuencia,

$$\mathbb{P}(V \neq W) = p - e^{-p}p = p(1 - e^{-p}) \leq p^2. \tag{38}$$

Usando la desigualdad (38) pueden obtenerse las siguientes cotas:

$$\sup_{k \geq 0} |\mathbb{P}(V = k) - \mathbb{P}(W = k)| \leq p^2, \tag{39}$$

$$\sum_k |\mathbb{P}(V = k) - \mathbb{P}(W = k)| \leq 2p^2. \tag{40}$$

La cota (39) se deduce de observar que

$$\begin{aligned}
|\mathbb{P}(V = k) - \mathbb{P}(W = k)| &= |\mathbb{E}[\mathbf{1}\{V = k\}] - \mathbb{E}[\mathbf{1}\{W = k\}]| \\
&= |\mathbb{E}[\mathbf{1}\{V = k\} - \mathbf{1}\{W = k\}]| \\
&\leq \mathbb{E}[|\mathbf{1}\{V = k\} - \mathbf{1}\{W = k\}|] \\
&\leq \mathbb{E}[\mathbf{1}\{V \neq W\}] \\
&= \mathbb{P}(V \neq W).
\end{aligned}$$

La cota (40) se deduce de observar que para todo $k = 0, 1, \dots$

$$\begin{aligned}
|\mathbb{P}(V = k) - \mathbb{P}(W = k)| &= |\mathbb{P}(V = k, W \neq k) - \mathbb{P}(W = k, V \neq k)| \\
&\leq \mathbb{P}(V = k, V \neq W) + \mathbb{P}(W = k, V \neq W),
\end{aligned}$$

y luego sumar sobre los posibles valores de k:

$$\sum_k |\mathbb{P}(V = k) - \mathbb{P}(W = k)| \leq 2\mathbb{P}(V \neq W).$$

$\square$

Nota Bene. Esta técnica, denominada técnica de acoplamiento de variables aleatorias, permite probar (sin usar la fórmula de Stirling) que la distribución Binomial converge a la distribución Poisson.

Teorema 2.5 (Le Cam)**.** *Sean $X_1, \dots, X_n$ variables aleatorias independientes con distribución Bernoulli de parámetros $p_1, \dots, p_n$, respectivamente y sea $S = \sum_{i=1}^n X_i$. Entonces*

$$\sum_k |\mathbb{P}(S = k) - \mathbb{P}(N = k)| \leq 2 \sum_{i=1}^n p_i^2, \tag{41}$$

donde N es una variable aleatoria con distribución Poisson de media $\lambda = \sum_{i=1}^n p_i$.

Demostración. Sean $U_1, \dots, U_n$ variables aleatorias independientes con distribución común $\mathcal{U}[0,1)$. Construimos variables aleatorias acopladas $V_i \sim$ Bernoulli(p_i) y $W_i \sim$Poisson(p_i), $i = 1, \dots, n$:

$$V_i := \mathbf{1}\{U_i \in I_1(p_i)\}, \qquad W_i := \sum_{k=0}^{\infty} k\mathbf{1}\{U_i \in J_k(p_i)\},$$

y las sumamos

$$S^* = \sum_{i=1}^n V_i, \qquad N = \sum_{i=1}^n W_i.$$

Por construcción, las variables $V_1, \dots, V_n$ son independientes y con distribución Bernoulli(p_i), respectivamente, y entonces, la variable S^* tiene la misma distribución que S; las variables $W_1, \dots, W_n$ son independientes y tienen distribución Poisson(p_i), respectivamente, y entonces, la variable N tiene distribución Poisson de media $\lambda = \sum_{i=1}^n p_i$.

Observando que cada k

$$|\mathbb{P}(S^* = k) - \mathbb{P}(N = k)| \leq \mathbb{P}(S^* = k, N \neq k) + \mathbb{P}(N = k, S^* \neq k).$$

se obtiene que

$$\sum_k |\mathbb{P}(S^* = k) - \mathbb{P}(N = k)| \leq 2\mathbb{P}(S^* \neq N).$$

Si $S^* \neq N$, entonces $V_i \neq W_i$ para algún $i = 1, \dots, n$. En consecuencia,

$$\mathbb{P}(S^* \neq N) \leq \sum_{i=1}^n \mathbb{P}(V_i \neq W_i) \leq \sum_{i=1}^n p_i^2.$$

$\square$

Corolario 2.6 (Aproximación Poisson)**.** *Para cada $k \geq 0$*

$$\lim_{n\to\infty} \binom{n}{k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \left(\frac{\lambda}{n}\right)^k = e^{-\lambda}\frac{\lambda^k}{k!}$$

Demostración. Sean $U_1, \dots, U_n$ variables aleatorias independientes con distribución común $\mathcal{U}[0,1)$. Para cada $i = 1, \dots, n$ definimos parejas de variables aleatorias (V_i, W_i) independientes

$$V_i := \mathbf{1}\{U_i \in I_1(p)\}, \qquad W_i := \sum_{k=0}^{\infty} k\mathbf{1}\{U_i \in J_k(p)\}.$$

Por construcción, $V_i \sim \text{Bernoulli}(p)$ y $W_i \sim \text{Poisson}(p)$, en consecuencia las sumas

$$S = \sum_{i=1}^{n} V_i, \qquad N = \sum_{i=1}^{n} W_i$$

son variables aleatorias con distribuciones Binomial(n,p) y Poisson(np), respectivamente. De acuerdo con la demostración del Teorema de Le Cam tenemos que

$$\left| \binom{n}{k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \left(\frac{\lambda}{n}\right)^{k} - e^{-\lambda}\frac{\lambda^k}{k!} \right| = |\mathbb{P}(S = k) - \mathbb{P}(N = k)| \le 2np^2 = 2\frac{\lambda^2}{n} \to 0.$$

$\square$

Teorema 2.7. *Supongamos que para cada n, $X_{n,1}, \dots, X_{n,r_n}$ son variables aleatorias independientes con distribución Bernoulli$(p_{n,k})$. Si*

$$\sum_{k=1}^{r_n} p_{n,k} \to \lambda \ge 0, \qquad \max_{1\le k\le r_n} p_{n,k} \to 0, \tag{42}$$

entonces

$$\mathbb{P}\left(\sum_{k=1}^{r_n} X_{n,k} = i\right) \to e^{-\lambda}\frac{\lambda^i}{i!}, \qquad i = 0, 1, 2, \dots. \tag{43}$$

Si $\lambda = 0$, el límite (43) se interpreta como 1 para $i = 0$ y 0 para $i \ge 1$. En el caso $r_n = n$ y $p_{n,k} = \lambda/n$, (43) es la aproximación Poisson a la binomial. Notar que si $\lambda > 0$, entonces (42) implica que $r_n \to \infty$.

Demostración. Sea $U_1, U_2, \dots$ una sucesión de variables aleatorias independientes, con distribución común $\mathcal{U}[0,1)$. Definimos

$$V_{n,k} := \mathbf{1}\{U_k \in I_1(p_{n,k})\}.$$

Las variables $V_{n,1}, \dots, V_{n,r_n}$ son independientes y con distribución Bernoulli$(p_{n,k})$. Puesto que $V_{n,1}, \dots, V_{n,r_n}$ tienen la misma distribución que $X_{n,1}, \dots, X_{n,r_n}$, (43) se obtiene mostrando que $V_n = \sum_{k=1}^{r_n} V_{n,k}$ satisface

$$\mathbb{P}(V_n = i) \to e^{-\lambda}\frac{\lambda^i}{i!}. \tag{44}$$

Ahora definimos

$$W_{n,k} := \sum_{i=0}^{\infty} i\mathbf{1}\{U_k \in J_i(p_{n,k})\}$$

$W_{n,k}$ tiene distribución Poisson de media $p_{n,k}$. Puesto que las $W_{n,k}$ son independientes, $W_n = \sum_{k=1}^{r_n} W_{n.k}$ tiene distribución Poisson de media $\lambda_n = \sum_{k=1}^{r_n} p_{n,k}$. De la desigualdad $1-p \leq e^{-p}$, se obtiene como consecuencia que

$$\begin{aligned}\mathbb{P}(V_{n,k} \neq W_{n,k}) &= \mathbb{P}(V_{n.k} = 1 \neq W_{n,k}) = \mathbb{P}(U_k \in I_1(p_{n,k}) - J_1(p_{n,k})) \\ &= p_{n,k} - e^{-p_{n,k}} p_{n,k} \leq p_{n,k}^2,\end{aligned}$$

y por (42)

$$\mathbb{P}(V_n \neq W_n) \leq \sum_{k=1}^{r_n} p_{n,k}^2 \leq \lambda_n \max_{1 \leq k \leq r_n} p_{n,k} \to 0.$$

(44) y (43) se obtienen de observar que

$$\mathbb{P}(W_n = i) = e^{-\lambda_n} \frac{\lambda_n^i}{n!} \to e^{-\lambda} \frac{\lambda^n}{n!}.$$

□

3. Cuentas con exponenciales

3.1. Motivación: pasaje de lo discreto a lo continuo

Para fijar ideas consideraremos una conversación telefónica y supondremos que su duración es un número entero de segundos. La duración de la conversación será tratada como una variable aleatoria T cuya distribución de probabilidades $p_n = \mathbb{P}(T = n)$ es conocida. La línea telefónica representa un sistema físico con dos estados posibles "ocupada" (E_0) y "libre" (E_1).

Imaginemos que cada segundo se decide si la conversación continúa o no por medio de una moneda cargada. En otras palabras, se realiza una sucesión de ensayos Bernoulli con probabilidad de éxito p a una tasa de un ensayo por segundo y se continúa hasta el primer éxito. La conversación termina cuando ocurre el primer éxito. En este caso la duración total de la conversación, el *tiempo de espera*, tiene distribución geométrica $p_n = (1-p)^{n-1}p$. Si en un instante cualquiera la línea está ocupada, la probabilidad que permanezca ocupada por más de un segundo es $(1-p)$, y la probabilidad de transición $E_0 \to E_1$ en el siguiente paso es p. En este caso esas probabilidades son independientes de cuánto tiempo estuvo ocupada la línea.

La descripción de los tiempos de espera mediante modelos discretos presupone la cuantización del tiempo y que los cambios solo pueden ocurrir en las épocas $\varepsilon, 2\varepsilon, \dots$. El tiempo de espera T más sencillo es el tiempo de espera hasta el primer éxito en una sucesión de ensayos Bernoulli con probabilidad de éxito $p(\varepsilon)$. En tal caso $\mathbb{P}(T > n\varepsilon) = (1 - p(\varepsilon))^n$ y el tiempo medio de espera es $\mathbb{E}[T] = \varepsilon/p(\varepsilon)$. Este modelo puede se puede refinar haciendo que ε sea cada vez más chico pero manteniendo fija la esperanza $\varepsilon/p(\varepsilon) = 1/\lambda$. Para un intervalo de duración t corresponden aproximadamente $n \approx t/\varepsilon$ ensayos, y entonces para ε pequeño

$$\mathbb{P}(T > t) \approx (1 - \lambda\varepsilon)^{t/\varepsilon} \approx e^{-\lambda t}. \tag{45}$$

Este modelo considera el tiempo de espera como una variable aleatoria discreta distribuida geométricamente y (45) dice que "en el límite" se obtiene una distribución exponencial.

Si no discretizamos el tiempo tenemos que tratar con variables aleatorias continuas. El rol de la distribución geométrica para los tiempos de espera lo ocupa la *distribución exponencial*. Es *la única variable continua dotada de una completa falta de memoria*. En otras palabras, la probabilidad de que una conversación que llegó hasta el tiempo t continúe más allá del tiempo $t+s$ es independiente de la duración pasada de la conversación si, y solo si, la probabilidad que la conversación dure por lo menos t unidades de tiempo está dada por una exponencial $e^{-\lambda t}$.

Nota Bene Si en un momento arbitrario t la línea está ocupada, entonces la probabilidad de un cambio de estado durante el próximo segundo depende de cuan larga ha sido la conversación. En otras palabras, *el pasado influye sobre el futuro*. Esta circunstancia es la fuente de muchas dificultades en problemas más complicados.

3.2. Distribución exponencial

Se dice que la variable aleatoria T tiene *distribución exponencial de intensidad* $\lambda > 0$ y se denota $T \sim \text{Exp}(\lambda)$ si la función de distribución de T es de la forma

$$F_T(t) := \mathbb{P}(T \leq t) = \left(1 - e^{-\lambda t}\right) \mathbf{1}\{t \geq 0\}. \tag{46}$$

En tal caso T admite la siguiente función densidad de probabilidades

$$f_T(t) = \lambda e^{-\lambda t} \mathbf{1}\{t \geq 0\}. \tag{47}$$

Media y Varianza. Los valores de la esperanza y la varianza de T son, respectivamente, $\mathbb{E}[T] = 1/\lambda$ y $\mathbb{V}(T) = 1/\lambda^2$.

3.3. Suma de exponenciales independientes de igual intensidad

Teorema 3.1. Sean $T_1, T_2, \dots, T_n$ variables aleatorias independientes, idénticamente distribuidas, con distribución exponencial de intensidad $\lambda > 0$. La suma $S_n = T_1 + \cdots + T_n$ admite una densidad de probabilidades de la forma

$$f_{S_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \mathbf{1}\{t > 0\} \tag{48}$$

y su función de distribución es

$$F_{S_n}(t) = \left(1 - e^{-\lambda t} \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{i!}\right) \mathbf{1}\{t \geq 0\}. \tag{49}$$

En otras palabras, la suma de n variables aleatorias independientes exponenciales de intensidad $\lambda > 0$ tiene distribución Gamma de parámetros n y λ: $\Gamma(n, \lambda)$.

Demostración. Por inducción. Para $n = 1$ no hay nada que probar: $S_1 = T_1 \sim \text{Exp}(\lambda)$. Supongamos ahora que la suma $S_n = T_1 + \cdots + T_n$ admite una densidad de la forma (48). Debido a que las variables aleatorias S_n y T_{n+1} son independientes, la densidad de $S_{n+1} = S_n + T_{n+1}$ se obtiene convolucionando las densidades de S_n y T_{n+1}:

$$\begin{aligned}
f_{S_{n+1}}(t) &= (f_{S_n} * f_{T_{n+1}})(t) = \int_0^t f_{S_n}(t-x) f_{T_{n+1}}(x)dx \\
&= \int_0^t \lambda e^{-\lambda(t-x)} \frac{(\lambda(t-x))^{n-1}}{(n-1)!} \lambda e^{-\lambda x} dx \\
&= \lambda e^{-\lambda t} \frac{\lambda^n}{(n-1)!} \int_0^t (t-x)^{n-1} dx = \lambda e^{-\lambda t} \frac{\lambda^n}{(n-1)!} \frac{t^n}{n} \\
&= \lambda e^{-\lambda t} \frac{(\lambda t)^n}{n!}.
\end{aligned}$$

Las funciones de distribución (49) se obtienen integrando las densidades (48). Sea $t \geq 0$, integrando por partes puede verse que

$$\begin{aligned}
F_{S_n}(t) &= \int_0^t f_{S_n}(s)ds = \int_0^t \frac{(\lambda s)^{n-1}}{(n-1)!} \lambda e^{-\lambda s} ds \\
&= -\frac{(\lambda s)^{n-1}}{(n-1)!} e^{-\lambda s} \bigg|_0^t + \int_0^t \frac{(\lambda s)^{n-2}}{(n-2)!} \lambda e^{-\lambda t} ds \\
&= -\frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} + F_{S_{n-1}}(t). \qquad (50)
\end{aligned}$$

Iterando (50) obtenemos (49). □

Nota Bene. En la demostración anterior se utilizó el siguiente resultado: si $T_1, \dots, T_n$ son variables aleatorias independientes, entonces funciones (medibles) de familias disjuntas de las T_i también son independientes. (Para más detalles ver el Capítulo 1 de Durrett, R., (1996). *Probability Theory and Examples*, Duxbury Press, New York.) □

3.4. Mínimos

Lema 3.2. Sean T_1 y T_2 dos variables aleatorias independientes y exponenciales de intensidades λ_1 y λ_2, respectivamente. Vale que

$$\mathbb{P}(T_1 < T_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}. \qquad (51)$$

Demostración. La probabilidad $\mathbb{P}(T_1 < T_2)$ puede calcularse condicionando sobre T_1:

$$\begin{aligned}
\mathbb{P}(T_1 < T_2) &= \int_0^\infty \mathbb{P}(T_1 < T_2 | T_1 = t) f_{T_1}(t)dt = \int_0^\infty \mathbb{P}(t < T_2)\lambda_1 e^{-\lambda_1 t} dt \\
&= \lambda_1 \int_0^\infty e^{-\lambda_2 t} e^{-\lambda_1 t} dt = \lambda_1 \int_0^\infty e^{-(\lambda_1+\lambda_2)t} dt = \frac{\lambda_1}{\lambda_1 + \lambda_2}.
\end{aligned}$$

□

Teorema 3.3. Sean $T_1, T_2, \dots, T_n$ variables aleatorias exponenciales independientes de intensidades $\lambda_1, \lambda_2, \dots, \lambda_n$, respectivamente. Sean T y J las variables aleatorias definidas por

$$T := \min_i T_i, \qquad J := \text{ índice que realiza } T.$$

Entonces, T tiene distribución exponencial de intensidad $\lambda_1 + \cdots + \lambda_n$ y

$$\mathbb{P}(J = j) = \frac{\lambda_j}{\lambda_1 + \cdots + \lambda_n}.$$

Más aún, las variables T y J son independientes.

Demostración. En primer lugar, hay que observar que $T > t$ si y solo si $T_i > t$ para todo $i = 1, \dots, n$. Como las variables $T_1, T_2, \dots, T_n$ son exponenciales independientes de intensidades $\lambda_1, \lambda_2, \dots \lambda_n$ tenemos que

$$\mathbb{P}(T > t) = \prod_{i=1}^{n} \mathbb{P}(T_i > t) = \prod_{i=1}^{n} e^{-\lambda_i t} = e^{-(\lambda_1 + \cdots + \lambda_n)t}.$$

Por lo tanto, T tiene distribución exponencial de intensidad $\lambda_1 + \cdots + \lambda_n$.

En segundo lugar hay que observar que $J = j$ si y solo si $T = T_j$. Por lo tanto,

$$\mathbb{P}(J = j) = \mathbb{P}(T_j = \min_i T_i) = \mathbb{P}(T_j < \min_{i \neq j} T_i) = \frac{\lambda_j}{\lambda_1 + \cdots + \lambda_n}.$$

La última igualdad se obtiene utilizando el Lema 3.2 pues las variables T_j y $\min_{i \neq j} T_i$ son independientes y exponenciales con intensidades λ_j y $\sum_{i \neq j} \lambda_i$, respectivamente.

Finalmente, si para cada j definimos $U_j = \min_{i \neq j} T_i$, tenemos que

$$\begin{aligned}
\mathbb{P}(J = j, T \geq t) &= \mathbb{P}(t \leq T_j < U_j) \\
&= \int_t^\infty \mathbb{P}(T_j < U_j | T_j = s) \lambda_j e^{-\lambda_j s} ds \\
&= \lambda_j \int_t^\infty \mathbb{P}(U_j > s) e^{-\lambda_j s} ds = \lambda_j \int_t^\infty e^{-(\sum_{i \neq j} \lambda_i)s} e^{-\lambda_j s} ds \\
&= \frac{\lambda_j}{\lambda_1 + \cdots + \lambda_n} \int_t^\infty (\lambda_1 + \cdots + \lambda_n) e^{-(\lambda_1 + \cdots + \lambda_n)s} ds \\
&= \frac{\lambda_j}{\lambda_1 + \cdots + \lambda_n} e^{-(\lambda_1 + \cdots + \lambda_n)t}.
\end{aligned}$$

Lo que completa la demostración. □

Ejercicios adicionales

4. Sean T_1 y T_2 variables aleatorias independientes exponenciales de intensidad 2. Sean $T_{(1)} = \min(T_1, T_2)$ y $T_{(2)} = \max(T_1, T_2)$. Hallar la esperanza y la varianza de $T_{(1)}$ y de $T_{(2)}$.

5. *Suma geométrica de exponenciales independientes.* Sean $T_1, T_2, \ldots$ variables aleatorias independientes idénticamente distribuidas con ley exponencial de intensidad λ. Se define $T = \sum_{i=1}^{N} T_i$, donde N es una variable aleatoria con distribución geométrica de parámetro p, independiente de las variables $T_1, T_2, \ldots$. Hallar la distribución de T. (*Sugerencia*: Utilizar la fórmula de probabilidad total condicionando a los posibles valores de N y el desarrollo en serie de Taylor de la función exponencial.)

---

4. Bibliografía consultada

Para redactar estas notas se consultaron los siguientes libros:

1. Billingsley, P.: Probability and measure. John Wiley & Sons, New York. (1986)
2. Durrett R.:Probability. Theory and Examples. Duxbury Press, Belmont. (1996)
3. Feller, W.: An introduction to Probability Theory and Its Applications. Vol. 1. John Wiley & Sons, New York. (1957)
4. Feller, W.: An introduction to Probability Theory and Its Applications. Vol. 2. John Wiley & Sons, New York. (1971)
5. Grimmett, G. R., Stirzaker, D. R.: Probability and Random Processes. Oxford University Press, New York. (2001)
6. Meester, R.: A Natural Introduction to Probability Theory. Birkhauser, Berlin. (2008).
7. Meyer, P. L.: Introductory Probability and Statistical Applications. Addison-Wesley, Massachusetts. (1972)
8. Ross, S. M: Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, San Diego. (2004)
9. Soong, T. T.: Fundamentals of Probability and Statistics for Engineers. John Wiley & Sons Ltd. (2004)